Technical Note: Pre-computing and Analysing LSH Bits for Atlas Neural Networks

Atlas LSH neural networks differ from conventional neural networks in that the binary LSH representation is an explicit geometric measurement of the input space. This creates an opportunity to analyse and improve the representation *before* network training.

The key observation is that each random hyperplane produces two related measurements:

- **Random projection:** $p = r \cdot x$
- **LSH bit:** $b = \text{sign}(p)$

The projection is a continuous geometric measurement, while the bit is its binary proposition ("which side of the hyperplane?"). Rather than using only the binary output, both levels of information can be exploited.

Pre-computing the LSH Representation

For many datasets, the complete set of LSH bits can be computed once before training. This allows the representation itself to be analysed independently of the learning algorithm.

The preprocessing stage becomes

Input vectors

Random projections

LSH bits

Representation analysis

Atlas neural network

Unlike conventional neural networks, the quality of the geometric representation can therefore be evaluated before any gradient descent or weight updates occur.

Using the Random Projections

The underlying projections contain significantly more information than the binary bits alone.

Useful statistics include:

- projection margin |p|
- projection variance
- projection distribution
- partition balance
- robustness under perturbation

Hyperplanes whose projections remain close to zero across the dataset produce unstable bits. Conversely, hyperplanes with consistently large margins produce highly reliable geometric measurements.

The projection therefore provides a natural confidence measure for each bit.

Using the LSH Bits

Once the binary codes have been generated, they can be analysed statistically.

Examples include:

- bit entropy
- bit persistence under input perturbations
- redundancy between bits
- mutual information
- Hamming neighbourhood statistics
- occupancy of LSH cells

These metrics identify bits that contribute little information or duplicate other measurements.

Stability Filtering

One particularly attractive preprocessing step is robustness analysis.

Small perturbations are applied to the input vectors and bit flip probabilities are measured.

Bits that frequently change under tiny perturbations are effectively measuring hyperplanes lying extremely close to the data manifold. Such bits are likely to be noisy and may be discarded or down-weighted.

In many cases, the projection margins alone may provide a computationally cheaper estimate of this instability.

Joint Projection and Bit Analysis

The greatest advantage comes from analysing the random projections and LSH bits together.

A hyperplane can be evaluated using both continuous and binary criteria:

- Is the partition balanced?
- Does it produce high entropy?
- Are the projection margins sufficiently large?
- Is the bit stable?
- Is the information redundant with other hyperplanes?

Rather than treating every random hyperplane as equally valuable, each becomes a geometric sensor whose quality can be measured objectively.

An Iterative Sensor Selection Procedure

Instead of accepting a single random collection of hyperplanes:

1. Generate a large pool of random projections.
2. Compute projections and LSH bits.
3. Analyse projection and bit statistics.
4. Remove poor-quality hyperplanes.
5. Generate replacements.
6. Repeat until the sensor array has desirable statistical properties.

This preserves the random nature of LSH while improving the overall quality of the representation.

Perspective

Viewed this way, the LSH layer is not merely a hashing mechanism but a large array of simple geometric sensors.

Each sensor asks a one-bit geometric question:

> Which side of this random hyperplane does the input lie on?

The associated random projection measures how confidently that question is answered.

This suggests that Atlas neural networks should treat the LSH layer as a measurable and optimisable sensing stage rather than as a fixed preprocessing step. By pre-computing the LSH representation and jointly analysing both the continuous projections and the binary bits, the network can be presented with a cleaner, more robust, and more informative geometric representation before learning begins.